DOCUMENT RESUME

ED 317 589 TM 014 647

AUTHOR Cushing, Katherine S.; And Others
TITLE Test Wise or Test Foolish: Effects of Riverside Materials on Test Taking Skill Instruction.
PUB DATE Mar 89
NOTE 22p.; Revision of a paper presented at the Annual Meeting of the National Council on Measurement in Education (San Francisco, CA, March 28-30, 1989).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Achievement Tests; Comparative Testing; *Elementary School Students; Grade 4; Grade 5; Instructional Effectiveness; *Instructional Materials; Intermediate Grades; *Standardized Tests; Student Characteristics; Test Coaching; Testing Programs; *Test Wiseness
IDENTIFIERS *Improving Test Taking Skills

ABSTRACT

The instructional effectiveness of commercially prepared test preparation materials was studied using Riverside Publishing Company's "Improving Test-Taking Skills" materials. The study further investigated differential effects of test-taking instruction as a result of student characteristics. In the first year of the study (1986-87), performance on standardized achievement tests of 182 fourth graders receiving an average of 12 hours and 212 fifth graders receiving an average of 9.25 hours of instruction with the Riverside materials was matched with that of students receiving test-taking instruction without the Riverside materials (223 fourth-graders, and 215 fifth-graders). In the replication year, students in nine schools received about 10 hours of Riverside instruction. Samples used in the second year included 129 fourth-graders and 92 fifth-graders receiving the Riverside materials, and 148 fourth-graders and 109 fifth-graders not receiving the Riverside materials. In the first year of the study, fourth-grade students receiving formal instruction in test taking did increase their scores, although teacher-made or commercial materials performed comparably well. In the second year, the Riverside method instruction appeared to have resulted in improved test scores only in mathematics for the fourth grade. No beneficial effects of test-taking instruction were found in the fifth grade in either year. No clear-cut patterns of differential effects for sex, socioeconomic status, or ethnicity were apparent for grade 4; results for grade 5 were similar. The policy implications of these findings are discussed. Two tables contain study results. (SLD)

Test Wise or Test Foolish:

Effects of Riverside Materials on

Test Taking Skill Instruction

Katherine S. Cushing
Harrison School District
Colorado

Darrell L. Sabers
University of Arizona

Judith A. Kerrins
Pueblo School District No. 60
Colorado

Revision of a paper presented at the Annual Meeting of the National Council on Measurement in Education, San Francisco, March, 1989. The authors extend appreciation to Riverside Publishing Company for cooperation in providing materials for this study.

Test-wise or Test Foolish:
Effects of Riverside Materials on
Test-taking Skill Instruction

Teaching test-taking skills appears to be a reasonable option for improving student achievement test scores. There is a widely held belief that student performance on standardized achievement tests is affected by construct irrelevant test variance, one type of which is called test-wiseness (Messick, 1989). Researchers studying test-wiseness have concluded that test-wiseness exists, that individuals possess different amounts of test-wiseness, and that test-wiseness can be taught (Callenbach, 1973; Gibb, 1964; Jongsma & Warshauer, 1975). Further, studies assessing the efficacy of teaching test-taking skills to students suggest that examinees who learn generic test-taking skills generally obtain higher scores on measures of achievement (Bangert-Drowns, Kulik, & Kulik, 1983; Samson, 1985). Because there is a need to maximize precision of test scores for accountability purposes and a political desire for "above average" achievement test scores, educational researchers, school district administrators, school board members, and professional organizations have advocated teaching test-taking skills to students (Downey, 1977; Ligon & Jones, 1981; Rawl, 1984).

There are, however, only a few studies available regarding the efficacy of commercially prepared materials for test-taking skill instruction, perhaps because these materials are relatively new on the market. Costars' 1980 study used Random House's Scoring High in Reading to teach fourth-grade students achievement test-taking behaviors. After two months of intervention no significant differences were found between treatment and control group students on the Metropolitan Achievement Test. In 1981 Crowe used Random House's Scoring High in Reading to teach experimental group students test-taking

skills daily for six weeks while control group students received additional instruction in mathematics. Statistical analysis revealed significant effects for fourth graders on five of seven subtests in mathematics and reading of the Comprehensive Tests of Basic Skills. However, no significant effects were found for the fifth graders who used the practice program in test-taking skills. Crowe suggested that since it was the third year the CTBS has been administered "students may have acquired as much sophistication in test-taking at this point as their developmental stage of thinking will permit" (1981, p. 88).

Deaton, Halpin, and Alford (1987) investigated performance on the California Achievement Tests (CAT) for students in first, second, fourth, and fifth grades who received instruction in test-taking skills using Scoring High on the CAT as compared to control group students who received no formal instruction in test-taking skills. Statistical analyses revealed some significant differences between the groups on some of the CAT subtests but, in general, the Scoring High program did not produce consistent increases in student scores.

A Chicago Public Schools study (Borger, Perlman, and van der Ploeg, 1987) compared the effectiveness of four test preparation programs (Random House's Scoring High materials, Riverside's Improving Test-Taking Skills, Continental Press' On Target for Tests, and Hammond's Reading Skills for Standardized Testing) in training students in test-taking skills and in assisting students in their abilities to generalize test-taking strategies thereby resulting in improved standardized test scores. Borger et al. acknowledge some internal validity concerns with this study so the findings must be considered with caution. While results indicated that students in all treatment groups showed greater improvement in their knowledge of test-taking skills compared to students in the control group and also showed an improved attitude toward testing after training in test-taking skills, the expected gain in achievement test scores for students in the treatment groups did not occur. No

statistically significant achievement gains were noted at any grade level. Borger et al. (1987) concluded that

> based on the results of this study, no recommendation can be made for school systems to purchase packaged instructional programs to teach test taking strategies. The packages did not produce results that were superior to whatever informal test-wiseness training was already in place. (pp. 33-34)

A 1987 study conducted by Benson-Pfiefle used Riverside's Improving Test-Taking Skills materials to teach test-taking skills to sixth grade students attending Seventh-Day Adventist schools. Significant differences were reported between the mean of the treatment group and the mean of the control group on the Visual (sic), Concepts, Problems, and Total Mathematics ITBS subtests, with no significant differences between boys and girls, and no differential benefit for low scoring and high scoring students. However, the group of subjects for this study was so unique (all students were enrolled in Seventh-Day Adventist Schools with an average student-teacher ratio of approximately 7:1; almost all students scored above the mean prior to any intervention; significant parental involvement and support) that the generalizability of the findings is severely limited.

Thus, the results of research using commercially prepared instructional materials is somewhat confusing. In general, studies of test-taking skills instruction suggest systematic instruction in such skills usually results in improved student achievement test scores (Bangert-Drowns, Kulik, & Kulik, 1983; Samson, 1985). However, the effectiveness of specific, commercially prepared instructional programs in improving student performance on standardized achievement tests remains in question. Commercially prepared programs may be effective, but little empirical evidence could be found to document such effectiveness. Deaton, et al. (1987) emphasized this point,

> More research is needed before educators can use with confidence --or not use at all-- the various intervention strategies that are available including the extensively commercially prepared programs. (p. 150)

A Study of Commercial Test Preparation Materials

In order to gather further information regarding the instructional effectiveness of commercially prepared materials a study was designed using Riverside's Improving Test-Taking Skills materials. This study further investigated any differential effects of test-taking skill instruction as a result of student demographic characteristics such as sex, SES level, ethnicity, or achievement level. Because this study used a large, intact, diverse data base with test-taking skill instruction delivered by regular classroom teachers and with district administered standardized achievement tests used as the outcome measure it is an important contribution to the literature on test-wiseness instruction. Further, in order to verify study findings, the study was replicated a second year with a smaller, more closely monitored sample, thus, two years of study data are included in this analysis.

Summary of Procedures

The subjects for this study were fourth and fifth grade students in 15 elementary schools in an urban school district located in the Rocky Mountain area. Schools in which staff volunteered to participate were matched and then each pair was matched to a third non-volunteering school. A coin was flipped and one member of the initial pair was assigned to the Riverside materials (RM) group, the matched pair to the teacher made materials (TM) group, and the third school was designated as a member of the Control group. During the first year of the study fourth-grade students in the RM group received an average of 12.0 hours and fifth-grade students received an average of 9.25 hours of test-taking skill instruction using the Riverside Improving Test-Taking Skills instructional

materials. Teachers in the TM group reported their students received approximately 8.5 hours of test-taking skill instruction, but no Riverside materials were used during that instruction. Teachers in the Control group schools reported that they were unable to estimate the amount of instruction their students received in test-taking skills (since these teachers were unaware of the study it was assumed these students received test-taking instruction to the same degree that it had been taught by these teachers in previous years using materials the teachers had made or purchased).

Standardized achievement tests were administered in April. Participants' 1987 obtained NCE scores were paired with their 1986 obtained NCE scores for purposes of data analysis. For fourth-grade students the pairing was an ITBS to SRA match across all three groups; for fifth grade students the pairing was SRA to SRA. Although the fourth grade match across tests (ITBS to SRA) was not ideal, because that match occurred for all fourth grade students in all three groups the effect should be the same regardless of group assignment. Further, since the Riverside materials did not attempt to match standardized achievement test questions with the instructional materials this ITBS to SRA match should be considered inconsequential to study findings. Gain scores were computed (i.e., 1987 NCE minus 1986 NCE for each student). It is important to note that negative gain represents less than a years' growth (normative comparisons) rather than an actual decline in achievement. The Analysis of Variance (ANOVA) procedure was used to identify statistical significance. Because differences might be masked by achievement level, all data were disaggregated and analyzed by group assignment across four achievement levels (lower quarter, middle two quarters, and top quarter based on 1986 Composite score). Finally, effect sizes were computed to determine the magnitude of effect that might be expected from implementing a similar test-taking skills instructional program.

During the replication year (1987-88) only nine elementary schools participated (three in each group). Nine schools were selected because this allowed for better monitoring by district staff who met regularly with RM group teachers to facilitate the use and delivery of the Riverside instructional materials. Teachers in the RM group delivered an average of 10 hours of test-taking skill instruction using the Riverside materials. Teachers in the TM and Control groups were unable to estimate the number of hours they spent in test-taking skill instruction, however TM group teachers reported that test-taking skill instruction had not received the same focus that it had the previous year. For data analysis, 1988 NCE scores were paired with 1987 scores, all SRA to SRA pairings.

Discussion of Findings

Fourth Grade Data

For the first year of the study an ANOVA of gain scores for fourth grade students indicated significant main effects for group assignment and for achievement levels for the Composite battery and for the Mathematics subtest. Effect sizes indicated moderate effects for students in both the RM and TM groups for the Composite battery and for the Mathematics subtest. No statistically significant differences were indicated for the Reading subtest. (See Table 1)

For the second year of the study gain score analysis indicated main effects for achievement level for the Composite battery and for the Reading subtest. In general, low achieving students gained more than high achieving students (regression effect) regardless of their group assignment (RM, TM, or Control group). Effect size calculations indicated the Middle-Low group on the Composite battery and the Middle-High group on the Reading subtest benefitted most from the intervention. On the Mathematics subtest a main effect was indicted for Group assignment and a Group by Level interaction effect (students

in the RM group outgained students in the other two groups at all but the highest achievement level). Again, effect size calculations indicated the intervention was most beneficial for middle scoring students.

Do student achievement test scores increase as a result of formal instruction using the Riverside materials? The answer remains elusive. It appears for the first year of the study the answer is a qualified 'yes", although students who received test-taking skill instruction using teacher-made or teacher collected materials performed comparably well. When considering both years of data, the RM instruction appears to have resulted in improved test scores on the Mathematics subtest only.

What is the explanation for this finding? Perhaps time was the important variable rather than the Riverside materials per se, since students received comparable hours of instruction in test-taking skills. Perhaps a John Henry Effect occurred: teachers in the TM were highly motivated to provide some kind of instruction in test-taking skills since they wanted to participate in the study but were not selected to do so. An interview with one of the teachers in the TM group supports this hypothesis. She reported that teachers in her school got together and developed a test preparation instructional program "since we were not allowed to participate" in the study (C. Avalos, personal communication, September 9, 1987). Another hypothesis is that a trade off of test-taking skill instruction for content instruction resulted in comparable scores on standardized measures of achievement. Further, those teacher made materials used for test-taking skill instruction would, most likely, more closely match the curriculum than commercially prepared materials. So even if students received a similar amount of time in direct content instruction and in direct test-taking skill instruction the curricular match would have been different depending on the materials used. Finally, it is also important to remember that the data were not analyzed

at the school or teacher level. Analysis at either the school or teacher level might have provided important explanations for the pattern identified above; however, due to the sensitive nature of this issue district personnel requested that the data not be analyzed at the school or teacher level. That request was honored.

Fifth Grade Data

For the first year of the study an ANOVA of gain scores indicated significant effects for group assignment and achievement level for the Composite battery. Effect sizes were greatest for students in the TM group. Of those who received the treatment intervention (test-taking skill instruction using the Riverside materials) effect size was greatest for low achieving students. For the Reading subtest significant effects were indicated only for achievement levels, with low achieving students generally experiencing the greatest gains. For the Mathematics subtest positive effects were reported for students in the TM group. Again, moderate effect sizes were indicated for low and middle-low achieving students in both the RM and TM groups.

For the second year of the study a main effect for group assignment was indicated for the Composite battery, with students in the Control group and the TM group outgaining students in the RM group. For the Reading subtest significant effects were indicated only for achievement levels, with the greatest gains made by low achieving students (regression effect). On the Mathematics subtest a main effect for group assignment was indicated, with students in the Control group making the greatest gains. These data are particularly interesting because most of these students participated in the study both years, thus, students in the RM group had two years of instruction in test-taking skills.

For fifth grade students the answer to the research question regarding the efficacy of test-taking skill instruction using the Riverside materials appears to be "no". Students

who received formal instruction in test-taking skills using the Riverside materials generally scored lower than students who were in the Control group and who received whatever was considered "normal" instruction in test-taking skills. Again, analysis at the building or teacher level might have provided an explanation for this finding, but at the district's request that analysis was not done.

Secondary Analysis of Data by SEX, SES, and ETHNICITY

Fourth Grade Data

At the fourth grade level, during both years of the study, lower achieving students almost always made greater gains than higher achieving students regardless of group membership, sex, socioeconomic status, or ethnicity--in part due to the regression effect. Effect size calculations indicated the treatment was, generally, most beneficial for middle-low achieving students. For the first year of the study sex effects were minimal, and moderated by group membership and content tested; during the replication year no main effects or interaction effects for sex were indicated. SES effects were also moderated by group assignment, achievement level, and content area during the first year of the study, though no pattern was apparent and no main effects or interaction effects for SES were indicated during the replication year. For both years of the study no differential effects were identified for students based on ethnicity.

Thus, instruction in test-taking skills, whether undertaken formally using commercially prepared materials or informally using whatever materials teachers made or gathered together, appears not to have evidenced any clear-cut pattern of differential effects on the basis of sex, socioeconomic status, or ethnicity for fourth grade students.

Fifth Grade Data

The fifth grade data are a bit more confusing. During the first year of the study the

students' sex, socioeconomic status, and ethnicity had a greater bearing on the results, although this was not the case during the replication year. For the first year, differential effects of test-taking skill instruction were identified on the basis of sex, but these differential effects were moderated, to some degree, by both content area and achievement level. For the second year of the study no main effects of interaction effects for sex were indicated. For the first year of the study SES effects were also moderated by content domain, group assignment, and achievement level. During the second year of the study no main effects or interaction effects were indicted for SES. Ethnicity also had a greater effect during the first year of the study than during the second year, but again these effects were moderated by achievement level and content domain. And again, during the second year of the study, only on the Mathematics subtest was ethnicity an important variable- the greatest gains were made by Hispanics in the control group and by whites in the TM group.

In general, the results of the fifth grade data analysis are similar to the results of fourth grade analysis. There is no identifiable pattern of differential effects on the basis of sex, SES, or ethnicity. When effects are indicated they appear to be moderated by content domain and achievement level.

The Issue of Reduced Variability

Part of the argument for teaching test-taking skills is to obtain a better estimate of "true score" by a reduction in variability due to differences in test-taking abilities. If variance in reported test scores is due to differences in test-taking abilities among subjects rather than true differences in domain knowledge then one would expect a reduction in the standard deviation for students who received test-taking skill instruction as compared to those who did not receive the instruction. Table 2 reports means and standard deviations for students involved in this study. No systematic differences in standard deviations were

identified, and one cannot conclude that teaching test-taking skills resulted in more accurate estimates of true content-domain knowledge.

Policy Implications

This study was somewhat unique in public education because volunteers were randomly assigned to the treatment group or to a control group, and a second control group was also included. Instruction in test-taking skills was delivered by regular classroom teachers (rather than by the experimenter) and the outcome measure was a nationally standardized achievement test. Thus, **the study design accurately reflects the world of practice for classroom teachers.** The study assessed differential effects of instruction in test-taking skills on the basis of sex, socioeconomic status, ethnicity (Hispanic or white), and previous history of low achievement level. Finally, the study was replicated a second year with a smaller sample to ensure monitoring and delivery of intervention.

Policy implications from the findings of this study focus on three major questions: (1) whether the benefits of teaching test-taking skills are great enough to make this a priority within the district instructional curriculum, (2) whether specific commercially prepared materials should be purchased for the instruction of test-taking skills, and (3) if test-taking skills are taught, whether specific instructional groupings on the basis of demographic information or prior achievement levels would be recommended as beneficial for student learning and achievement.

Of course, the answers to these policy questions are not clear-cut, and, further, these policy issues are fraught with ethical, political, and economic overtones.

The answer to the first policy question, whether the benefits of teaching test-taking skills are great enough to make this a priority within the district instructional curriculum, is a qualified "possibly". Previous studies have suggested small but significant gains in

achievement test scores as a result of test-taking skill instruction. However, this study could not document such gains. At issue here might be the amount of time in which content instruction is lost because test-taking skill instruction is provided. It appears, from this study, that time lost on content instruction may not be compensated for by knowledge of how to go about taking a test. And in the long run time off task may result in less content knowledge by students. This should be considered carefully before instruction in test-taking skills is advocated. Matter (1986) addressed a similar issue when he wrote, "test preparation activities should not be additional activities imposed upon teachers. Rather they should be incorporated into the regular, ongoing instructional activities whenever possible" (p. 10).

If test-taking skill instruction is adopted, should specific instructional materials be purchased? That question is difficult to answer and probably cannot be answered by this study. The materials used in this study, Riverside's Improving Test-Taking Skills, were no more effective than teacher-made and teacher- gathered materials, assuming the effect of teaching ability, energy, enthusiasm, and knowledge were randomly distributed across all three groups. Further, according to Mehrens and Kaminski (1989) these materials would be considered "more ethical" than a program such as Random Houses' Scoring High series; yet, studies cited earlier in this paper suggest that Scoring High materials did not result in significantly improved student achievement test scores either (Borger et al., 1987; Costar, 1980; Crowe, 1981; Deaton et al., 1987). Because some students in this study benefitted from whatever materials teachers in the TM made or gathered, it might be useful to review these materials to be certain they are all appropriate instructional materials. Perhaps these materials were more directly matched to academic content, and therefore time spent learning and practicing test-taking skills was, at the same time, reinforcing grade-level

curriculum. The economic savings of using teacher-made or teacher-collected materials should be considered; however, any fiscal savings must be considered in light of teacher-hours spent making, collecting, or reproducing similar materials. If this time is taken from (or better spent in) planning and preparing for content instruction the cost savings may not be great enough to compensate for a predicted loss in content instruction and coverage. Finally, there is the issue of teacher energy, enthusiasm, or commitment. A dollar amount cannot be assigned to these variables. Will teacher commitment be greater or less if the materials are purchased? Will teacher energy be greater or less if they participate in developing instructional materials? Will teacher commitment be sustained or short-term if they are expected to make or gather together the materials and integrate them into the curriculum?

Ethical and political issues are at the center of the third policy question: whether specific instructional grouping on the basis of demographic information or prior achievement levels would be recommended for test-taking skill instruction. The findings from this study hint at possible instructional grouping patterns for test-taking skill instruction: instructional grouping assignments could be made on the basis of ability level. In general middle-low achieving students benefitted most from test-taking skill instruction, although the instruction was not delivered in segregated groups. Would low achieving students make even greater gains if they received intensive instruction in test-taking skills designed specifically for them?

The issue of ability grouping for instruction was addressed recently by Slavin (1987) who reported that

> ability grouping is maximally effective when done for only one or two subjects, with students remaining in heterogeneous classes most of the day; when it

> reduces student heterogeneity in a specific skill, when group assignments are frequently reassessed; and when teachers vary the level and pace of instruction according to students' needs. (p. 293)

Since middle-low achieving students appeared to benefit most from this instruction it would appear desirable to provide similar instruction for those students. If students are already ability grouped for at least some instruction (a typical practice in many elementary schools [Slavin, 1987]) perhaps test-taking skill instruction can be integrated into the regular curriculum throughout the school year without the need for additional ability grouping.

Instructional grouping on the basis of demographic features (sex, socioeconomic status, or ethnicity) is both politically and ethically questionable. Issues of providing equal opportunity for learning and equal access to instruction would likely emerge if or when such grouping practices became known. If instructional groups were treated differently in terms of content coverage such grouping might increase, rather than decrease, achievement test score discrepancy between groups. Further, the effects on self-concept and socialization skills are unknown, but many would argue they would be negative. Finally, because this is one of the first studies, and only one study, which assessed differential effects of instruction on the basis of s x, socioeconomic status, and ethnicity and because differential effects on the basis of such demographic characteristics were moderated by content domain further research is needed before any instructional grouping recommendations can be made with confidence.

Recommendations and Implications

Clearly, further studies are needed to document effectiveness in using commercially prepared materials for teaching test-taking skills. Additional studies which utilized different or multiple standardized achievement tests would increase the generalizability of study

findings. Better monitoring of classroom instruction to assess the degree to which the materials are being used and to assess the degree of test-taking skill instruction in control schools should be included in future studies as well. Such monitoring would increase the confidence with which educators could make instructional and policy recommendations.

Since results were different for fourth and fifth grade students in this study it is also possible that the test-curriculum match across the grade levels moderated the effects of the test-taking skill instruction. Further investigation of this issue in future test-wiseness studies might be valuable. Or, as Crowe (1981) suggested, since 1987 was the third year and 1988 the fourth year of SRA achievement test administration in this district, students may have been saturated with test familiarity. Therefore, the findings may depend on students' familiarity with the dependent variable.

Although all these questions need to be answered, we urge caution in the adoption of a test-taking skills instructional curriculum for the improvement of student achievement test scores. We recognize the political and accountability issues involved in improving student achievement test scores; however, the results of this study suggest test-taking skill instruction using generic materials not based on regular class content will not improve scores significantly. If reporting high standardized achievement test scores is the goal, other intervention measures (e.g., Honig, 1990) should be considered.

REFERENCES

Bangert-Drowns, R. L., Kulik, J. A., & Kulik, C-L. C. (1983). Effects of coaching programs on achievement test performance. Review of Educational Research, 53, 571-585.

Benson-Pfiefle, B. L. (1987). Effects on achievement test scores resulting from teaching test-taking skills in the fifth grade. (Doctoral Dissertation, Loma Linda University, 1987). University Microfilms International, Order No. 8723525.

Borger, J. B., Perlman, C., & van der Ploeg, A. J. (1987, April). A comparison of the effectiveness of four test preparation programs. Paper presented at the Annual Meeting of the National Council on Measurement in Education, Washington, D. C.

Callenbach, C. (1973). The effects of instruction and practice in content-independent test-taking techniques upon the standardized reading test scores of selected second-grade students. Journal of Educational Measurement, 10, 25-30.

Cohen, S. A., & Foreman, D. I. (1978). Scoring high in reading. New York: Random House School Division.

Costar, E. (1980). Scoring high in reading: The effectiveness of teaching achievement test-taking behaviors. Elementary School Guidance and Counseling, 15, 157-159.

Crowe, D. E. (1981). The use of practice programs to improve testscores of elementary school students. (Doctoral dissertation, University of South Carolina, 1981). Dissertation Abstracts International, 42, 3116A. (University Microfilms No. 81-29, 451).

Deaton, W. L., Halpin, G., & Alford, T. (1987). Coaching effects on California Achievement Test scores in elementary grades. Journal of Educational Research, 80, 149-155.

Downey, G. W. (1977). Is it time we started teaching children how to take tests? American School Board Journal, 164, 27-30.

Gibb, B. G. (1964). Test-wiseness as Secondary Cue Responses. (Doctoral Dissertation - Stanford University) Ann Arbor, Michigan: University Microfilms, No. 64-7643.

Honig, B. (1990, February 28). 'Comprehensive' strategy can improve schools. Education Week, pp. 56, 31.

Jongsma, E. A. & Warshauer, E. (1975). The effects of instruction in test-taking skills upon student performance on standardized achievement tests. Baton Rouge, LA: Louisiana State University. (ERIC Document Reproduction Service No. ED 114 408)

Ligon, G., & Jones, P. (1981). Preparing students for standardized testing: One district's perspective. Austin, TX: Austin Independent School District. ERIC Document No. 218 319.

Mehrens, W. A. & Kaminski, J. (1989). Methods for Improving standardized test scores: Fruitful, fruitless or fraudulent? Educational Measurement: Issues and Practices, 8, 14-22.

Messick, S. (1989). Validity. In R. L. Linn (ed.)., Educational Measurement (3rd ed.). Washington, D.C.: National Council on Measurement in Education/ American Council on Education.

Pritchard, F. (1983). Improving Test-Taking Skills. Chicago: Riverside.

Rawl, E. H. (January, 1984). Test-taking strategies can be the key to improving test scores. NASSP Bulletin, 108-112.

Samson, G. E. (1985). Effects of training in test-taking skills on achievement test performance: A quantitative synthesis. Journal of Educational Research, 78, 261-266.

Slavin, R. E. (1987). Ability grouping and student achievement in elementary schools: A best evidence synthesis. Review of Educational Research, 57, 293-336.

Table 1
Effect Sizes Based on NCE Gains[1]

Fourth Grade Students	1986 TO 1987			1987 TO 1988		
	ES1	ES2	ES3	ES1	ES2	ES3
Composite Battery						
Group Total	.30	.32	-.01	-.01	.06	-.05
Low	.04	.32	-.29	-.18	-.27	.07
Mid-Low	.53	.56	-.04	.43	.29	.19
Mid-High	.34	.16	.24	-.14	-.26	.09
High	.32	.24	-.08	-.38	-.42	.03
Reading Subtest						
Group Total	.24	.17	.07	-.25	-.06	-.16
Low	.04	.02	.02	-.43	-.19	-.25
Mid-Low	.24	.37	-.11	-.18	.02	-.11
Mid-High	.41	.13	.29	.00	.41	-.35
High	.24	.16	.08	-.35	-.38	.03
Mathematics Subtest						
Group Total	.25	.33	-.10	.34	-.04	.37
Low	-.15	.28	-.49	.25	-.15	.40
Mid-Low	.62	.61	.02	.98	.50	.60
Mid-High	.37	.25	.17	.35	-.13	.35
High	.16	.19	-.02	-.22	-.41	.16
Fifth Grade Students						
Composite Battery						
Group Total	-.09	.25	-.25	-.36	-.01	-.27
Low	.18	.16	.02	.18	-.21	.02
Mid-Low	.04	.12	-.06	-.15	.05	-.18
Mid-High	-.07	.45	-.31	-.19	.22	-.34
High	-.45	.31	-.60	-1.22	-.52	-.47
Reading Subtest						
Group Total	-.12	-.15	.03	.03	.00	.03
Low	-.06	-.05	-.01	.44	-.12	.33
Mid-Low	-.20	-.34	.11	.06	-.06	.09
Mid-High	.07	-.14	.24	-.05	.22	-.29
High	-.34	-.12	-.17	-.16	-.13	-.03
Mathematics Subtest						
Group Total	.05	.28	-.22	-.50	-.05	-.36
Low	.22	.27	-.06	-.11	.24	-.27
Mid-Low	.22	.31	-.10	-.27	.13	-.56
Mid-High	.00	.45	-.32	-.47	-.12	-.20
High	-.24	.16	-.36	-1.31	-.57	-.49

ES1 = (Mean of RM group - Mean of control group) / SD control group ES2 = (Mean of TM group - Mean of control group) / SD control group
ES3 = (Mean of RM group - Mean of TM group) / SD TM control group
[1] Gain Score = 1987 NCE minus 1986 NCE or 1988 NCE minus 1987 NCE

Table 2
Means (and Standard Deviations) of
SRA Achievement Test Scores
for all Participant Groups

First Year of the Study

Fourth Grade

	RM n = 182	TM n = 223	CONTROL n = 216
Composite			
1986	52.56 (17.95)	50.34 (17.45)	49.66 (20.08)
1987	56.41 (19.78)	54.33 (17.24)	50.38 (20.81)
Reading			
1986	49.90 (17.88)	49.25 (16.86)	49.08 (19.84)
1987	54.20 (19.61)	52.71 (17.95)	50.54 (21.01)
Mathematics			
1986	50.85 (18.42)	48.35 (17.84)	47.27 (19.35)
1987	57.28 (20.68)	55.95 (17.66)	49.95 (19.76)

Fifth Grade

	RM n = 212	TM n = 215	CONTROL n = 242
Composite			
1986	57.09 (16.68)	53.95 (17.78)	53.03 (17.45)
1987	56.67 (16.01)	56.46 (20.73)	53.38 (18.30)
Reading			
1986	58.18 (15.68)	54.99 (18.96)	54.10 (18.33)
1987	57.60 (16.58)	54.08 (19.29)	54.80 (17.93)
Mathematics			
1986	56.36 (18.38)	54.65 (18.12)	51.90 (18.06)
1987	56.84 (18.37)	58.22 (20.60)	51.69 (19.73)

Second Year of the Study

Fourth Grade

	RM n = 129	TM n = 148	CONTROL n = 133
Composite			
1987	58.42 (19.63)	57.85 (19.53)	55.41 (15.94)
1988	58.10 (18.95)	56.06 (19.56)	54.17 (16.96)
Reading			
1987	58.96 (20.66)	57.28 (20.59)	54.33 (18.10)
1988	55.30 (19.22)	55.57 (18.74)	53.38 (16.81)
Mathematics			
1987	57.84 (18.31)	57.90 (18.30)	57.35 (14.64)
1988	60.93 (18.45)	56.17 (19.50)	56.15 (18.08)

Fifth Grade

	RM n = 92	TM n = 109	CONTROL n = 101
Composite			
1986	52.53 (17.80)	52.07 (14.73)	49.71 (18.73)
1987	59.68 (20.18)	56.71 (14.82)	50.66 (17.53)
1988	59.47 (18.94)	59.69 (17.75)	53.70 (20.67)
Reading			
1986	49.54 (18.49)	49.71 (15.44)	49.99 (18.12)
1987	57.87 (19.88)	56.03 (15.90)	50.86 (18.04)
1988	59.70 (17.58)	57.51 (16.58)	52.39 (18.21)
Mathematics			
1986	51.37 (18.90)	50.58 (16.42)	47.52 (19.10)
1987	61.25 (20.61)	58.27 (15.91)	51.29 (16.75)
1988	60.03 (20.90)	62.31 (18.41)	55.88 (20.79)